Developing Distributed Collaboration Systems at NASA: A Report from the Field

Irma Becerra-Fernandez*, Helen Stewart** and Chris Knight**
*Florida International University, Decision Sciences and Information Systems, Miami, FL
**NASA-Ames Research Center, Moffett Field, CA

1. Introduction

Web-based collaborative systems have assumed a pivotal role in the information systems development arena. While business to customers (B-to-C) and business to business (B-to-B) electronic commerce systems, search engines, and chat sites are the focus of attention; web-based systems span the gamut of information systems that were traditionally confined to internal organizational client server networks. For example, the *Domino* Application Server allows *Lotus Notes* ™ [1] users to build collaborative intranet applications and *mySAP.com* ™ [2] enables web portals and e-commerce applications for *SAP* users. This paper presents the experiences in the development of one such system: *Postdoc*, a government off-the-shelf web-based collaborative environment. Issues related to the design of web-based collaborative information systems, including lessons learned from the development and deployment of the system as well as measured performance, are presented in this paper. Finally, the limitations of the implementation approach as well as future plans are presented as well.

2. Postdoc: a web-based environment to enable group collaboration in NASA

One tool frequently touted under the auspices of knowledge management (KM) is document management. At the core of a document management system is a centralized repository, an electronic storage medium with a primary storage location that affords multiple access points. The document management system essentially organizes and stores information. A document management system unifies an aggregate of relevant information conveniently in one location through a common interface. Document management builds upon the central repository by adding support to the classification and organization of information, unifying the actions of storage and retrieval of documents instituted over a platform independent system. The document management collaborative application increases communication, thus allowing the sharing of organizational knowledge.

At present, NASA has many dispersed information databases in place with powerful and sophisticated infrastructures supporting some information management (IM). It is nonetheless being improved as NASA s functionally and geographically diverse organizations work toward enhanced collaboration, information/technology fusion, and ultimately a transition from IM to KM. A bridge between IM and KM, Intelligent Information Management (IIM), can be constructed as an extension of the current information technology (IT) infrastructure. The goal of IIM is to facilitate the creation of knowledge from the information resources within the various enterprise program and mission operations, such as logistics of deployments, program strategies, to defining missions. This can be done by enhancing the use of the information by human beings, allowing them to make more knowledgeable decisions, or by enabling automated systems to refine knowledge from our information resources for use in performance support or intelligent training systems.

NASA took its first step in IIM with the New Millennium Documentation Systems success (Becerra-Fernandez et. al., 2000); which resulted in the development of a new application, Postdoc[3], that would ensure that all of NASA could easily use such a system Agency-wide. The innovation began back in 1994 with the New Millennium Program. The program was large and had a large resource issue. Goldin[4] was championing the faster, better, cheaper, smarter interplanetary missions. The program began with six technology teams, three Earth Orbiting missions, and six Deep Space missions. Already the program was beginning to experience cuts from budget and needed different resources to ensure mission objectives and

---

[1] www.lotusnotes.com

[2] www.sap.com

[3] www.postdoc.arc.nasa.gov

[4] Daniel S. Goldin, NASA administrator 1992-2001.

schedules. Postdoc made this possible through a new foundation of tools, making the application source more portable, and adding features ensuring the users that they had full control of their data anytime, anywhere, easily. The Postdoc software required a development effort of five person-years of software coding and testing.

Postdoc is a multi-user, web-based document management system that is currently used primarily for document storage and retrieval, including word processing documents, spreadsheets, slides, illustrations, images, video, audio, software archives, and others. Unlike a traditional web site where the webmasters alone construct it, a Postdoc web site is "easily" constructed by its users who can login to the site to add, delete, and organize documents tailored to their choice. Users need not have any specialized web site construction knowledge or skills such as HTML or programming. Users create documents on their own computer using many popular applications (such as Microsoft¤ Word, Excel, and PowerPoint) then easily upload them to the Postdoc server. The server can then automatically convert the document s format (for example into PDF), enabling any user to view the document even without the creating application. By virtue of being a web-based application, Postdoc can be used by anyone, anywhere. Postdoc is available for free for government use, and NASA-ARC supports the deployment of this innovative tool, not only within NASA centers but also at other government agencies.

NASA s various program use of the Postdoc software sparked an awareness in the agency on how large programs can create virtual workspaces to be shared among geographically distributed teams. Postdoc provides a portal for the agency to begin greater information collaboration with research and technology partners who are geographically dispersed nationally and internationally.

Postdoc is a completely different application from other commercial-off-the-shelf (COTS) software primarily in its ease of use, insuring accessibility, and allowing users the availability to connect and grant access to industrial and academia partners in the information sharing spaces. Other differences in the Postdoc software system compared to software similarly developed (i.e., NMP-Doc. Sys., Livelink, Virtual Research Center, TechDocs, etc.) are:

1. Postdoc allows users to invite, authorize, and grant access to potential team members that are geographically dispersed. Postdoc allows users to define the types of access to the information , allowing users to manage their own information spaces.
2. Postdoc has document versioning, checkout, and revision control, cut and paste options for organization.
3. There is event logging available to the users, auto-conversion of file-type to portable document format (pdf), and an on-demand document conversion to a broader option of document types (i.e., postscript, html, text, and digital assistant file types).
4. Postdoc has utilities making it fax capable, a fully featured email list manager, and a virtual collaborative platform independent file system.
5. Postdoc has online help integrated into the system. For maintenance of the software system, there are administration features for system administrators monitoring the translations of the documents, maintaining statistics of events, and more. There is also event-logging and reporting available to all users.

3. Results and Lessons Learned from the Postdoc Implementation at NASA

Postdoc (NASA-Case Number ARC 15008, 2000), which has been used ARC and cross the Agency since 1997, has proven to be a very effective tool for organizing, storing, and retrieving documents of all types (e.g., word-processing, spreadsheet, presentation, image, video, audio, executable).° Currently Postoc has 3779 users in total, and supports approximately 30 NASA programs spanning across the whole agency (e.g., Advanced Technologies and Mission Studies, Aeronautics Design and Test Environment (at ARC), Space Shuttle Independent Assessment, Aviation Safety (at Goddard Space Flight Center), ESDIS Advanced Technology/Systems (at ARC), SMA Build II (for collaboration between the FAA and NASA). Thanks to the flexibility in the system that allows users to define their own collaborative structure and control of access, the agency users have found efficient ways to share data in the system with geographically dispersed partners. Several Postdoc servers have been installed at other NASA Centers, (e.g., Dryden Flight Center, Langley Research Center, Johnson Space Center). Postdoc is in use by a significant portion of the Federal Government (i.e., Department of Defense, National Institutes of Standards and Technology, Naval Research Laboratories, and National Imagery and Mapping Agency).

Many programs have used the system to conduct document sharing of all types, such as review of proposals responding to NASA Research Announcements, program reporting and tracking, and scientific reporting.

Many users have done call for papers involving large conferences (i.e., AAAI conferences, IJCAI, and many other professional society activities). Several centers use the system for working on program initiatives, collaborating on technology and mission plans, proposals, distributing information throughout the centers and more. The Astrobiology Institute (containing an additional 500 users in a separate server from the main ARC server with 3,000 users internationally distributed) uses the system to begin publication sharing, data sharing, sharing science results (picture library) and conferencing (in conjunction with whiteboard and video conferencing). ARC ISO Certification Directorate uses the software as a documentation system for storage of ISO documentation, correction action requests, and revisions. Field experiments successfully used the system to share results in real-time with geographically dispersed teams. For example, Postdoc provided a largely automated Web site according to specifications, for the collection and dissemination of the MARS rover data, to display such data on the Web with elegant point-and-click accessibility. The support and use of Postdoc, allowed the Marsokhod mission scientists to share and collaborate on science results easily, which was extremely valuable. Many programs and initiatives have reported that the software allows them to communicate smoothly across platforms, across software versions, and collaborate effectively, nationally and internationally.

It s estimated that the current use of the software based in the Computational Sciences Division has saved the agency over $4M a year in data sharing and travel costs. Over 25 NASA programs cross-enterprise have saved at least $100K and up to $200K in expenditures; of travel, tedious email issues with large documents, or having to buy commercial software, by using the software system provided by the Postdoc Project in the Computational Sciences Division. The agency continues to save as the user-base on the current server continues to grow significantly.

4. The next generation in Distributed Collaboration: the Complex Object Relationship Engine (CORE) Project

Postdoc has been very effective in meeting the targeted scope of a document management system, and continues to be used today to manage thousands of documents across many divisions of NASA. However, Postdoc's ability to support KM is limited by this initial scope (Becerra-Fernandez et. al., 2000). A plan is in place to re-architect the Postdoc system with a larger, more generic scope, in order to meet KM needs, as well as the growing needs of other projects throughout NASA that are looking to expand Postdoc's functionality into (often drastically) different applications. Currently, Postdoc's architecture is being re-designed as part of the Complex Object Relationship Engine (CORE) project, a next generation multi-application object environment for collaboration. The software teams re-designing Postdoc are using an object-oriented infrastructure in support of application plug-in extension capability to accomplish two goals. First, an object-oriented (OO) architecture facilitates the objectives of allowing user-managed data collection through object attributes and application extensions (Knight and Aha, 2000). This will allow Postdoc s user base to easily integrate software extensions that address their specific needs, thus permitting them to test potential science and KM enhancements. Second, this new architecture will allow CORE to be able to support the addition of user application tools, or plug-ins. CORE Design goals include the ability to handle metadata and support workflow application logic with a focus on KM. With these changes CORE will be much more flexible and be able to support a KM environment, such as user-definition of workflow scripts. The ability to create modular extensions to be shared among different groups, will yield a library of modules that can benefit each project. Adequate access control and authentication is also being re-designed in order to support a wide variety of users. The CORE project will also include the development of a DARWIN component. The DARWIN goal is to streamline the information flow for U.S. Aerospace Industry's use of NASA technology and resources in their aircraft/spacecraft design cycle and development processes. This is of high significance; it fosters cooperation and synergy with NASA's Information Technology Base program and the new Intelligent Synthesis Environments Initiative, which is aimed towards NASA's engineering, science culture and creative process.

5. The Technologies to Implement CORE

The new CORE incorporates workflow management functionality as one of its modules. The addition of workflow capabilities is important in the context of tracking changes in the states of documents, proposals drafted, submitted, reviewed, voted on, and tallied in a virtual collaborative environment (Becerra-Fernandez et. al., 2001). Dispersed groups will benefit from a true dynamic, collaborative environment. Workflow functionality provides the CORE user with the following features and capabilities:

1. Graphical Workflow Design: The graphical user interface (GUI) that provides the means for creating workflow processes by defining the flow and the tasks that need to be performed.
1. Rules: The logic that needs to be integrated into the workflow definition.
2. Exception Handling: The system must be able to handle exceptions and errors.
3. Monitoring: The ability to monitor the status of workflow instances.
4. Auditing: Involves being able to log and report time and cost of the workflow process.
5. Notification: The workflow application must be able to inform users of new tasks, warn them of late tasks, and reroute tasks to other users.
6. Document attachments: The ability to effectively attach documents to the workflow.

The most important objective for the workflow functionality is that it seamlessly integrate with the CORE system. While the features that are to be incorporated into this workflow management system are not vastly different or superior to workflow tools that are currently on the market, the real benefit comes from how it augments the current capabilities of Postdoc. Other tools, such as Staffware2000, FileNet s Visual Workflow, and IBM s MQ series, offer similar features such as visual workflow design, user defined rules and application logic, web functionality, and role-oriented workflow task assignments. The CORE workflow management system, however, is specifically designed to operate within the CORE system architecture. This allows the workflow functionality to access and manage all object-oriented features of this system including access control, document handling, document metadata management, and user application components.

CORE implements off-the-shelf technologies as much as possible, including support to:
1. Database back-end built upon relational (Oracle) and object (ODBMS) databases. Object information is transparently mapped to any off-the-shelf RDBMS/ODBMS for scalability and for efficient searching and indexing.
2. CORBA for multi-process, multi-system shared access to the CORE server.
3. WebDAV, web protocol extensions for remote document management.
4. Xalan — an XML transformation engine for converting application-generated XML, into a client-viewable form.
5. JAVA and JDBC — used for high-performance and reliability when serving and interfacing to databases.
6. Python — an optional client-side programming for ease-of-enhancement at runtime.

6. Conclusions and Directions for Future Research

Although many gains have been achieved from the organizational use of distributed collaboration systems such as Postdoc, a limiting factor to distributed collaboration is the lack of connectivity among disparate collaboration systems in the organization. Currently, a number of distributed collaboration systems like Postdoc have been deployed throughout NASA. For example, while Postdoc s deployment spans several NASA Centers, it s use is concentrated at ARC. Similarly, TechDocs, a similar web-based collaboration system was developed and deployed at Kennedy Space Center, Virtual Research Center was developed and deployed at Marshall Space Flight Center, and COTS software such as Livelink was deployed throughout Lewis Space Flight Center. Such proliferation of collaboration systems is due to in part to a tendency towards the not-invented-here syndrome, personal preferences, lack of awareness about activities at other Centers, and the lack of trust sometimes prevalent in large organizations.

A major objective for CORE is to enhance the usefulness of distributed collaboration systems like the ones aforementioned by providing needed enterprise-wide connectivity amongst these systems. This is of high significance; as it will foster cooperation and synergy amongst the many programs, including the NASA's Information Technology Base program, the Intelligent Synthesis Environments, and the Design for Safety Program Inititiative which is aimed towards NASA's engineering, mitigating risk in a systems full life-cycle specifically the design process.° NASA needs to demonstrate an increase by a factor of ten over baseline, in percentage of participants in enterprise science and engineering design, development, and management activities using diverse geographically distributed electronic collaboration applications. This motivates the need for software tools that can be used to enhance communication among distributed project members and the information sources that they require. Postdoc began addressing this need; it is used by many, to share documents (e.g., proposals, schedules, budgets, revisions) with user-specified access settings (e.g., budgets are typically accessible to only a subset of a project team s members).° In addition, grouped users can conduct threaded email discussions that record decision-making, information sharing, and other project activities.° However, Postdoc and other NASA collaborative Environments are not yet capable to assist with other types of tasks (e.g., to help design experiments, or serve as a portal/conduit for relevant information sources), and are therefore limited in the direct ability to support scientific investigations and project engineering applications.° CORE s object-oriented architecture and intelligent software modules are designed to support distributed scientific and engineering collaborations.° In particular, the intension is to evolve these collaborative environments from successful document management tools into ones that can support advanced KM processes, capable of capturing and disbursing project knowledge among distributed members, and providing a platform for science instrument data integration.

An additional impact of this approach is to further enable the management of technical, scientific, and engineering data.° This coupling between the multiple data sources, both textual and numerical, will greatly enhance the collaborative capability of NASA s science and engineering community.° Given that knowledge management pilots and activities such as document management systems (like PostDoc, TechDocs, or LiveLink) are already in use by other agencies and thousands of users, we anticipate that this enhancement would prove invaluable to many key design and science experiments and data analysis environments (e.g., Aerospace, REVCON Program, Joint Strike Fighter Program, SOFIA Program, and NASA s Astrobiology Institute) that are currently using it to collaborate and share documents and scientific data.°

References

Becerra-Fernandez I., Stewart H., Del Alto M., Knight C (2000). Developing an Advanced Environment for Collaborative Computing. Proceedings of the Twelfth Annual International Florida Artificial Intelligence Research Symposium (FLAIRS), Orlando, Florida.

Becerra-Fernandez I., McCarthy K., Rodriguez J. (2001) An Infrastructure for Managing Knowledge using Intelligent Workflow. Proceedings of the Fourteenth Annual International Florida Artificial Intelligence Research Symposium (FLAIRS), Key West, Florida. Forthcoming

Knight, C., Aha, D. (2000). A Common Knowledge Framework and Lessons Learned. Proceedings of the AAAI Workshop on Intelligent Lessons Learned Systems, Austin, Texas.

NASA-Case Number ARC 15008 (2000) Inventions and Contributions Board Space Act Award Application, Title: Postdoc Software, April , 2000.